Classification of European mtDNAs From an Analysis of Three European Populations

Antonio Torroni,*,† Kirsi Huoponen,* Paolo Francalacci,‡ Maurizio Petrozzi,† Laura Morelli,‡ Rosaria Scozzari,†,§ Domenica Obinu,‡ Marja-Liisa Savontaus and Douglas C. Wallace***

**Department of Genetics and Molecular Medicine, Emory University School of Medicine, Atlanta, Georgia 30322, †Dipartimento di Genetica e Biologia Molecolare, Universita' di Roma "La Sapienza," 00185 Rome, Italy, ‡Istituto di Antropologia, Universita' di Sassari, 07100 Sassari, Italy, §Centro di Genetica Evoluzionistica, C.N.R., Rome, Italy and **Department of Medical Genetics, University of Turku, 20520 Turku, Finland*

Manuscript received March 12, 1996
Accepted for publication August 26, 1996

ABSTRACT

Mitochondrial DNA (mtDNA) sequence variation was examined in Finns, Swedes and Tuscans by PCR amplification and restriction analysis. About 99% of the mtDNAs were subsumed within 10 mtDNA haplogroups (H, I, J, K, M, T, U, V, W, and X) suggesting that the identified haplogroups could encompass virtually all European mtDNAs. Because both hypervariable segments of the mtDNA control region were previously sequenced in the Tuscan samples, the mtDNA haplogroups and control region sequences could be compared. Using a combination of haplogroup-specific restriction site changes and control region nucleotide substitutions, the distribution of the haplogroups was surveyed through the published restriction site polymorphism and control region sequence data of Caucasoids. This supported the conclusion that most haplogroups observed in Europe are Caucasoid-specific, and that at least some of them occur at varying frequencies in different Caucasoid populations. The classification of almost all European mtDNA variation in a number of well defined haplogroups could provide additional insights about the origin and relationships of Caucasoid populations and the process of human colonization of Europe, and is valuable for the definition of the role played by mtDNA backgrounds in the expression of pathological mtDNA mutations

TWO aspects of the mitochondrial DNA (mtDNA) make it particularly useful in human evolutionary studies. First, it is maternally transmitted and thus does not undergo recombination (GILES *et al.* 1980). Second, the mtDNA sequence evolution rate is much higher than that of the average nuclear gene (MIYATA *et al.* 1982; WALLACE *et al.* 1987). Consequently, a substantial number of mtDNA mutations have accumulated sequentially along radiating maternal lineages that have diverged as human populations colonized different geographical regions of the world. Restriction fragment length polymorphism (RFLP) studies of mtDNAs from a wide range of human populations have revealed a number of stable polymorphic sites in the mtDNA coding regions. These define related groups of mtDNAs called haplogroups. Most of the mutations observed in both mtDNA coding and control regions in modern human populations have occurred on these preexisting haplogroups and define the individual mtDNA types or haplotypes (TORRONI *et al.* 1993a; GRAVEN *et al.* 1995).

The majority of haplogroups have been shown to be continent-specific. In Africa, haplogroup L encompasses between 70 and 100% of the sub-Saharan mtDNAs (CHEN *et al.* 1995; GRAVEN *et al.* 1995). In Asia, ~55% of East Asian and Siberian mtDNAs are members of haplogroup M (BALLINGER *et al.* 1992; TORRONI *et al.* 1993b,c; CHEN *et al.* 1995; WALLACE 1995). Haplogroup M is subdivided into smaller subhaplogroups designated C, D, G and E. Most of the remaining Asian mtDNAs are encompassed by haplogroups A, B and F (TORRONI *et al.* 1994c). Among Native Americans, only four Asian haplogroups (A, B, C and D) are observed, thus indicating that these haplogroups predated the colonization of the Americas (SCHURR *et al.* 1990; WARD *et al.* 1991; HORAI *et al.* 1993; TORRONI *et al.* 1993a, 1994a,d).

A recent high resolution RFLP study of individuals of European ancestry living in the United States and Canada has revealed four European-specific haplogroups (H, I, J and K) that encompassed ~64% of the subjects analyzed (TORRONI *et al.* 1994b). While useful in delineating European mtDNA haplogroups, the diverse European origins of the samples precluded identification of associations between particular populations and haplogroups. Since such associations have provided powerful tools for reconstructing African, Asian, and Native American origins, we have endeavored to determine the nature and frequency of mtDNA variation in well defined European populations.

In this study, we report the extent and nature of mtDNA variation in three European populations, the

Corresponding author: Antonio Torroni, Dipartimento di Genetica e Biologia Molecolare, Universita' di Roma "La Sapienza," P.le A. Moro 5, 00185 Rome, Italy. E-mail: torroni@axcasp.caspur.it

Finns, the Swedes and the Tuscans of Central Italy. These populations encompass a large portion of the linguistic and genetic differentiation of modern European populations (CAVALLI-SFORZA and PIAZZA 1993; CAVALLI-SFORZA *et al.* 1994). The Finnish language belongs to the Uralic family, and the ancestral Finns probably were a mixture of preexisting populations of unknown origin and population groups of Baltic and Germanic ancestry (DE LA CHAPELLE 1993). The Swedish language belongs to the Germanic branch of the Indo-European family, and the ancestors of modern Swedes could have been either Neolithic farmers who expanded into Europe from the Middle East ~10,000 yr before present (YBP) (AMMERMANN and CAVALLI-SFORZA 1984; CAVALLI-SFORZA *et al.* 1988; SOKAL *et al.* 1991) or pastoral nomads that began to expand from the Eurasian regions north of the Black Sea ~6500 YBP (GIMBUTAS 1991). The Tuscan sample is from an area of Central Italy that was the core of the Etruscan colonization. The Etruscan language, although not completely known, is not of Indo-European origin and it was spoken in the area up to the first century A.D., when it was replaced by Latin, a language of the Indo-European family.

This survey revealed five new mtDNA haplogroups, in addition to the previous four described in Europeans from North America and one already reported in Asians. When combined, these 10 haplogroups encompassed virtually all Finnish, Swedish and Tuscan mtDNAs. This finding suggests that the same 10 haplogroups could encompass almost all mtDNAs in Europe and implies a scenario similar to that described in the Americas, where four haplogroups were found to encompass almost the entire mtDNA variation of Native American populations.

MATERIALS AND METHODS

Subjects: Of the 86 Scandinavian subjects analyzed, 49 were unrelated Finns collected from various regions of Finland and 37 were unrelated Swedes collected among the medical students of the Gothenburg University. DNAs from these samples were extracted from buffy coats using standard procedures. The DNAs from 48 unrelated Tuscan individuals whose maternal grandmothers were also born in Tuscany (Central Italy) were extracted from hair roots using the method described in VIGILANT *et al.* (1989). Sequence analysis of the two hypervariable segments of the mtDNA control region from these Tuscan samples was previously performed, and it is described in FRANCALACCI *et al.* (1996). All samples (1–52) analyzed by FRANCALACCI and collaborators (1996), except samples 6, 8, 52 and 50, were also analyzed in this study. Samples 6, 8 and 52 were maternally related to other individuals of the sample and were included as sequence controls by FRANCALACCI *et al.* (1996). Sample 50 was discarded because of intersample contamination.

mtDNA analysis: The entire mtDNA of each Finn and Swedish sample was amplified in nine overlapping fragments by PCR using the same primer pairs and amplification conditions described in TORRONI *et al.* (1992, 1993a). Each of the nine PCR segments was then digested with 14 restriction endonucleases (*Alu*I, *Ava*II, *Bam*HI, *Dde*I, *Hae*II, *Hae*III, *Hha*I, *Hinc*II, *Hin*fI, *Hpa*I, *Msp*I, *Mbo*I, *Rsa*I, *Taq*I) and tested for the A to G sequence polymorphism at nucleotide position (np) 12308 in the $tRNA^{Leu}$ gene, previously observed in European mitochondrial disease patients (LAUBER *et al.* 1991). This mutation was detected by PCR using a mismatched primer with the 3′ end adjacent to np 12308 and a G at np 12312. When the 12308G mutation is present, this generates a *Hin*fI site. In addition, all subjects were screened for the presence of a *Bst*NI site at np 13704, an *Acc*I site at np 15254, and a *Nla*III site at np 4577. Restriction fragments were resolved through electrophoresis in NuSieve plus SeaKem agarose (FMC BioProducts) gels and visualized by UV-induced fluorescence. The resulting haplotypes are reported in APPENDIX A.

The phylogenetic relationships among and between Finnish and Swedish haplotypes were inferred by parsimony analysis. The dendrograms were rooted using the African haplotype AF71 ("African outgroup"). This haplotype, which has been observed in the Senegalese, is a member of the African-specific haplogroup L (CHEN *et al.* 1995) and has been previously used as an outgroup in TORRONI *et al.* (1993a,b, 1994a–d). Maximum parsimony (MP) trees were generated through random addition of sequences using the Tree Bisection and Reconnection (TBR) algorithm (PAUP 3.1.1, SWOFFORD 1993). The PAUP analysis was terminated at 350 trees after 30,000 replications. These trees were obtained by saving no more than 10 MP trees for each replication. Although shorter trees could exist, none were observed in our analyses. A 50% majority-rule consensus of the 350 MP trees generated by the TBR algorithm was also obtained. A NJ tree was obtained from *P* distances by using the program MEGA 1.01 (KUMAR *et al.* 1993). Currently available program packages, including MEGA 1.01 and PHYLIP 3.5c (FELSENSTEIN 1993), do not allow the users to compute genetic distances from restriction site data, but only from nucleotide sequences. To obtain *P* distances, we recoded the 0/1 haplotype matrix by replacing the 0's with A's, and the 1's with G's, and treated the restriction site data set as a sequence data set. The reliability of haplogroups observed in the NJ tree was assessed by bootstrap analysis (500 replications). Intrahaplogroup divergence estimates from restriction analysis data were calculated using the maximum likelihood procedure of NEI and TAJIMA (1983) and the mtDNA evolution rate of 2.2–2.9% per million yr (TORRONI *et al.* 1994d).

In contrast to the Finnish and Swedish samples, the limited amount and poor quality of the DNA extracted from hair roots of the Tuscans did not allow a complete haplotype analysis of their mtDNAs. However, it was possible to amplify 12 short PCR fragments (~200–300 bp each) from each Tuscan mtDNA (Table 1), and to screen these fragments for the presence of 14 polymorphic restriction sites that define the mtDNA haplogroups (TORRONI *et al.* 1994b,c; CHEN *et al.* 1995).

RESULTS

Haplotype analysis of Finnish and Swedish mtDNAs: Fifty-two haplotypes (E1–E52), defined by 84 polymorphic restriction sites, were observed among the 86 Finnish and Swedish samples analyzed (Table 2, APPENDIX A). Only 13 haplotypes were observed in more than one subject, and only five (E2, E11, E19, E24, E45) were shared by Finns and Swedes. Of these haplotypes E2, E19 and E45 were the most common haplotypes, representing 15.1, 7.0 and 8.1%, respectively, of the total samples analyzed.

TABLE 1

Primers used for partial PCR amplification of Tuscan mtDNAs

Haplogroup	Polymorphic sites[a]	Primer coordinates[b]	Segment sizes (bp)
H	−7025 *Alu*I	6890–6909, 7131–7115	242
I	−1715 *Dde*I	1615–1643, 1894–1874	280
	+8249 *Ava*II	8188–8207, 8366–8345	179
	+10028 *Alu*I	9911–9932, 10107–10088	197
J	−13704 *Bst*OI	13583–13605, 13843–13824	261
K	−9052 *Hae*II	8829–8845, 9184–9163	356
	+12308 *Hin*fI	12104–12124, 12338–12309[c]	235
T	+13366 *Bam*HI	13172–13190, 13403–13384	232
	+15606 *Alu*I	15409–15428, 15701–15682	293
U	+12308 *Hin*fI	12104–12124, 12338–12309[c]	235
V	−4577 *Nla*III	4308–4325, 4739–4720	432
W	+8249 *Ava*II	8188–8207, 8366–8345	179
	−8994 *Hae*III	8829–8845, 9184–9163	356
X	−1715 *Dde*I	1615–1643, 1894–1874	280
L (African)[d]	+3592 *Hpa*I	3388–3408, 3717–3701	330
M (Asian)[e]	+10397 *Alu*I	10270–10290, 10579–10557	309

[a] All Tuscan samples were also screened for the *Dde*I np 10394 site by using the same primer pair employed to test for the presence of the *Alu*I np 10397 site.

[b] Primers are numbered according to ANDERSON *et al.* (1981) with the 5′ → 3′ coordinates before the comma corresponding to the forward primers and those after the comma to the reverse primers.

[c] The reverse primer of this pair of oligonucleotides was altered in order to screen for an A to G transition at np 12308 in the $tRNA^{Leu}$ gene.

[d] CHEN *et al.* (1995).

[e] BALLINGER *et al.* (1992); TORRONI *et al.* (1994c); CHEN *et al.* (1995).

Comparison of polymorphic sites revealed that only a limited number of these sites were shared by different Scandinavian haplotypes, and that the shared polymorphic sites were usually observed in well defined and alternative associations (Table 3). This observation suggests that these polymorphisms are relatively ancient and preceded most of the additional variation accumulated on the Scandinavian mtDNAs. These specific sets of associated polymorphisms were used to segregate the Scandinavian mtDNAs into groupings (Table 3). A first subdivision of Scandinavian mtDNAs is due to the presence or absence of a *Dde*I site at np 10394. Because the same dichotomization is also observed in mtDNAs from African (CHEN *et al.* 1995), Asian (BALLINGER *et al.* 1992; TORRONI *et al.* 1994c) and Native American (TORRONI *et al.* 1993a) populations, it has been proposed that the 10394 *Dde*I polymorphism is very ancient and predated the radiation of modern human populations (WALLACE 1995). Additional polymorphic sites subdivided Scandinavian mtDNAs into 10 smaller groups of related haplotypes (haplogroups H, I, J, K, M, T, U, V, W, and X) that encompassed ~99% of the Scandinavian mtDNAs (Table 3). Four of the haplogroups (H, I, J and K) had been described in detail among Europeans from North America (TORRONI *et al.* 1994b); one (haplogroup M) is the most common haplogroup in Asians (BALLINGER *et al.* 1992; TORRONI *et al.* 1994c; CHEN *et al.* 1995), while the remaining five haplogroups (T, U, V, W and X) were identified in this study. The frequencies of these haplogroups in our Finnish and Swedish samples and those previously reported in Europeans from North America are illustrated in Table 4.

Haplotype E39, observed in one Finnish subject, was the only Scandinavian haplotype that did not cluster in any of the haplogroups (Table 2). Haplotype E39 lacked the *Dde*I site at np 10394 (Table 3), placing it together with haplogroups H, T, U, V, W and X; but was devoid of other distinguishing polymorphisms. Thus it could not be placed in any of these haplogroups.

Parsimony and NJ analyses were performed to verify the reliability of the 10 haplogroups suggested by haplotype data. Parsimony analysis confirmed the clustering of Scandinavian mtDNAs into 10 haplogroups (Figure 1), each defined by the sets of associated polymorphisms illustrated in Table 3. The identity and topology of these haplogroups were also preserved in the 50% majority rule consensus of 350 MP trees (Figure 1), supporting the robustness of the haplogroups. However, the consensus tree did not retain the major dichotomization between haplogroups harboring (haplogroups I, J, K and M) or lacking (haplogroups H, V, T, U, W, and X) the 10394 *Dde*I site. The NJ tree further supported the subdivisions into the 10 haplogroups even though nodal bootstrap values could be obtained only for those haplogroups harboring more than one haplotype (Figure 2). In addition, the NJ tree confirmed the tendency to aggregate haplogroups according to the presence or absence of the 10394 *Dde*I

TABLE 2

MtDNA haplotypes in Finns and Swedes

Haplotype	Haplogroup	Population		N
		Finn	Swede	
E1	H	2	—	2
E2	H	9	4	13
E3	H	—	2	2
E4	H	1	—	1
E5	H	1	—	1
E6	H	1	—	1
E7	H	1	—	1
E8	H	1	—	1
E9	H	—	1	1
E10	H	—	1	1
E11	H	2	1	3
E12	H	1	—	1
E13	H	—	1	1
E14	H	—	1	1
E15	H	—	2	2
E16	H	—	1	1
E17	H	—	1	1
E18	H	1	—	1
E19	T	2	4	6
E20	T	1	—	1
E21	T	—	1	1
E22	T	—	1	1
E23	T	—	2	2
E24	U	1	1	2
E25	U	—	1	1
E26	U	1	—	1
E27	U	1	—	1
E28	U	2	—	2
E29	U	1	—	1
E30	U	1	—	1
E31	U	—	1	1
E32	U	1	—	1
E33	U	—	1	1
E34	U	—	1	1
E35	U	—	1	1
E36	V	—	1	1
E37	V	—	1	1
E38	V	1	—	1
E39	Other	1	—	1
E40	V	1	—	1
E41	W	2	—	2
E42	X	1	—	1
E43	X	1	—	1
E44	I	1	—	1
E45	K	2	5	7
E46	J	2	—	2
E47	J	2	—	2
E48	J	—	1	1
E49	J	1	—	1
E50	J	1	—	1
E51	J	1	—	1
E52	M (Asian)	1	—	1
Total		49	37	86

site. The exception is represented by haplogroup I, which is connected to haplogroup W with a nodal bootstrap *P* value of 40%. If it is confirmed, this aggregation could indicate that the 10394 *Dde*I site has undergone a certain number of parallelisms/reversals during recent human evolution. The nodal *P* values for haplogroups T, J, V and X were comprised between 90 and 63%. These were followed by haplogroups H and U with *P* values of 33 and 21%, respectively. Bootstrap values are commonly taken as the level of statistical confidence for the subset of haplotypes taken into consideration, and are considered statistically significant when $P \geq 95\%$. According to this, none of the *P* values observed in this European data set would appear statistically significant. However, it has to be considered that 95% bootstrap values have not been reported for a major clade in any of the numerous studies that have been dealing with human mtDNA variation [as examples, see HEDGES *et al.* (1991) and SOODYALL *et al.* (1996)]. This is because of homoplasy in conjunction with the relatively low number of stable and informative characters (BANDELT *et al.* 1995) and the fact that bootstrap values tend to underestimate true values, especially when many haplotypes or sequences are analyzed (LI and ZHARKIKH 1995).

In contrast to the parsimony trees, the NJ tree also suggests that haplogroups I, W and X are more closely related to each other than to the other haplogroups (Figure 2). The tendency to aggregate them in a major branch of the NJ tree is probably due to the fact that haplogroups I and X share a restriction site loss (*Dde*I np 1715), and haplogroups I and W share a site gain (*Ava*II np 8249) (Table 3).

Screening of Tuscan mDNAs for haplogroup-specific mutations: To determine if the 10 haplogroups found in the Finnish and Swedish mtDNAs were also characteristic of southern European populations, we screened 48 Tuscan mtDNAs for the 14 polymorphic sites listed in Table 5. These sites included most of the polymorphisms that define each of the haplogroups observed in Scandinavia, as well as the *Hpa*I np 3592 site that defines the African-specific haplogroup L (CHEN *et al.* 1995).

This analysis revealed that 98% of the Tuscan mtDNAs could be aggregated into the 10 haplogroups (Table 5). The *Alu*I np 7025 site loss in association with the lack of the *Dde*I np 10394 site was found in 41.7% of the Tuscan mtDNAs, thus indicating that haplogroup H is also the most common haplogroup in southern Europe (Table 4). The *Bam*HI np 13366 and *Alu*I np 15606 sites, which characterize haplogroup T, were observed in 10.4% of the Tuscan mtDNA. The *Hin*fI np 12308 site without the *Hae*II np 9052 site loss and the *Dde*I np 10394 site gain, which is typical of haplogroup U, was also observed in 10.4% of the Tuscans. The associated *Ava*II np 8249 site gain/*Hae*III np 8994 site loss, which defines haplogroup W, was observed in 2.1% of the Tuscans. The *Dde*I np 1715 site loss in the absence of the *Dde*I site at np 10394, *Alu*I site at np 10028, and *Ava*II site at np 8249 defines haplogroup X, and was observed

TABLE 3

Haplogroup-specific polymorphic sites in Finnish and Swedish mtDNAs

Sites[a]																		
−1715 *Dde*I	−4529 *Hae*II	−4577 *Nla*III	−7025 *Alu*I	+8249 *Ava*II	−8994 *Hae*III	−9052 *Hae*II	+10028 *Alu*I	+10394 *Dde*I	+10397 *Alu*I	+12308 *Hin*fI	+13366 *Bam*HI	−13704 *Bst*NI	+15606 *Alu*I	−15925 *Msp*I	−16065 *Hin*fI	+16389 *Bam*HI	Frequency (%)[b]	Haplogroup
−	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	40.7	H
+	+	−	−	+	−	−	+	+	−	−	−	−	−	−	−	+	1.2	I
−	−	−	−	−	−	−	−	+	−	−	−	+	−	−	+	−	9.3	J
−	−	−	−	−	−	+	−	+	−	+	−	−	−	−	−	−	8.1	K
−	−	−	−	−	−	−	−	−	−	−	+	−	+	+	−	−	12.8	T
−	−	−	−	−	−	−	−	−	−	+	−	−	−	−	−	−	17.3	U
−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	3.6	V
−	−	−	−	+	+	−	−	−	−	−	−	−	−	−	−	−	2.3	W
+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	2.3	X
−	−	−	−	−	−	−	−	+	+	−	−	−	−	−	−	−	1.2	M
−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	1.2	Other

[a] The polymorphic restriction sites are listed as either absent (−) or present (+) relative to the reference sequence (ANDERSON *et al.* 1981). A plus or minus indicates whether the listed haplogroups have that polymorphism.
[b] Finns + Swedes.

in 8.3% of the subjects. The *Dde*I np 1715 site loss in association with the *Dde*I site at np 10394, *Ava*II site at np 8249 and *Alu*I site at np 10028 defines haplogroup I, and was found in 4.2% of the subjects. The associated *Hin*fI np 12308 site gain and *Hae*II 9052 site loss, together with the *Dde*I site gain at np 10394 define haplogroup K and was found in 6.3% of the sample. Finally, the *Bst*NI np 13704 site loss in association with the *Dde*I np 10394 site that defines haplogroup J was observed in 14.6% of the Tuscans, with one of the seven haplogroup J mtDNAs (sample #21) also having the *Alu*I site gain at np 15606 site gain that is characteristic of haplogroup T mtDNAs (Table 5). The presence of the *Alu*I site at np 15606 in a haplotype J background suggests that this polymorphism arose as an independent mutational event. Neither the *Nla*III np 4577 site loss in the absence of the *Dde*I np 10394 site that defines haplogroup V, nor the *Alu*I np 10397 site in the presence of the *Dde*I np 10394 site that define haplogroup M were observed in any of the Tuscan mtDNAs (Table 5).

DISCUSSION

Ten haplogroups encompass almost all mtDNAs from three European populations: The mtDNA analysis of Swedish and Finnish samples revealed that virtu-

TABLE 4

Distribution of mtDNA haplogroups in three European populations

Haplogroups	Finland[a]	Sweden[a]	Tuscany[a]	North America[b]
H	20 (40.8)	15 (40.5)	20 (41.7)	40.0
I	1 (2.0)	0	2 (4.2)	7.4
J	7 (14.3)	1 (2.7)	7 (14.6)	9.1
K	2 (4.1)	5 (13.5)	3 (6.3)	7.4
M	1 (2.0)	0	0	0
T	3 (6.1)	8 (21.6)	5 (10.4)	ND
U	8 (16.3)	6 (16.2)	5 (10.4)	ND
V	2 (4.1)	2 (5.4)	0	ND
W	2 (4.1)	0	1 (2.1)	ND
X	2 (4.1)	0	4 (8.3)	ND
Others	1 (2.0)	0	1 (2.1)	ND
Total	49	37	48	

[a] Values are number of subjects with percentage in parentheses.
[b] Data from TORRONI *et al.* (1994b); ND, not determined. At the time of this study only mtDNA haplogroups H, I, J, K and M were defined. Values are percentage of subjects.

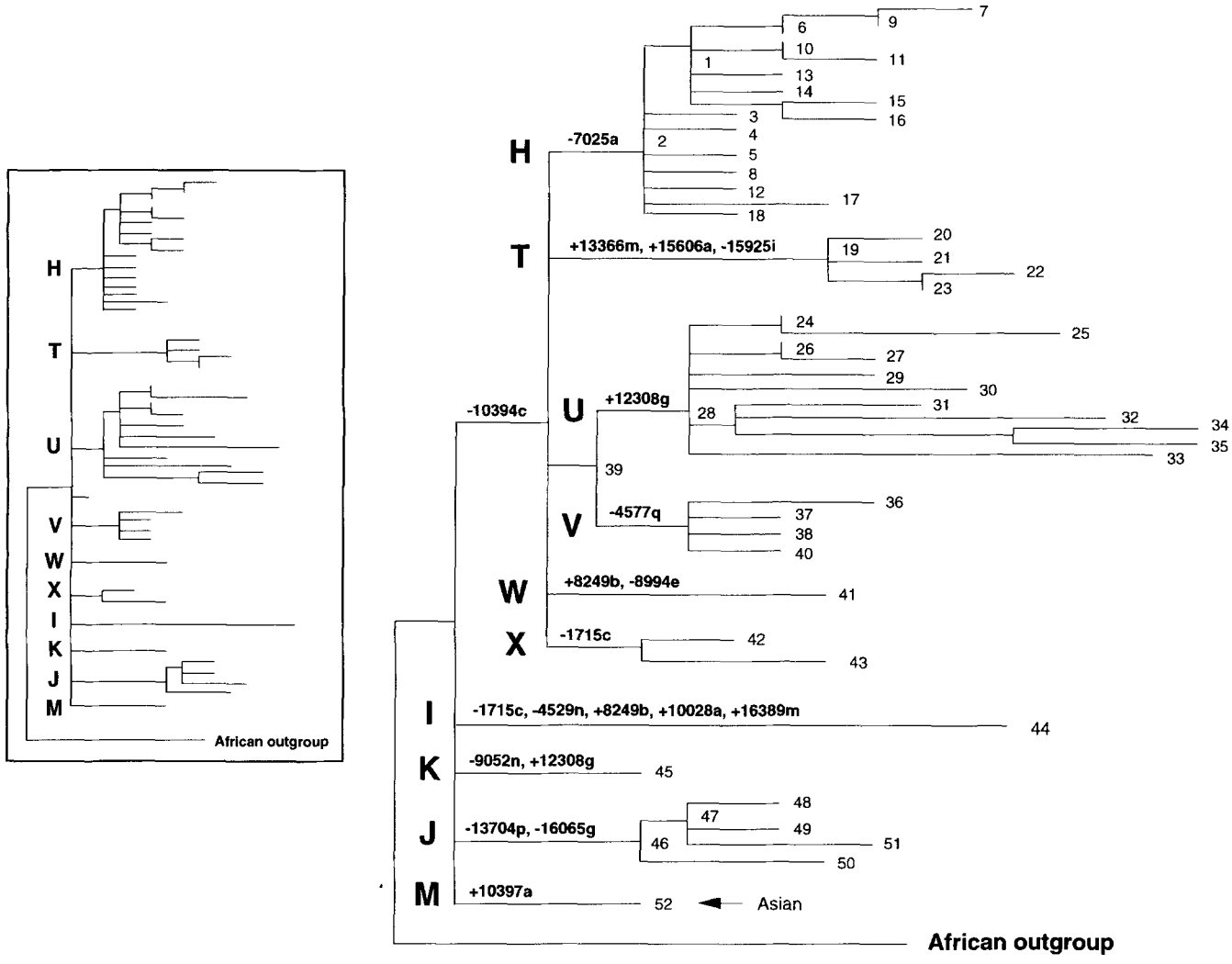

FIGURE 1.—Parsimony tree of Finnish and Swedish mtDNA haplotypes. This MP tree has been obtained by PAUP analysis and includes the 52 haplotypes (E1–E52) observed in 49 Finns and 37 Swedes. The numbers at the nodes or end of each branch indicate different mtDNA haplotypes. The tree was rooted by using the African haplotype AF71 ("African outgroup"). This haplotype is a member of the African-specific haplogroup L (CHEN *et al.* 1995) and has been previously used as an outgroup in TORRONI *et al.* (1993a,b; 1994a–d). The capital letters H, I, J, K, M, T, U, V, W and X indicate the 10 haplogroups, and the numbers associated with the lowercase letters indicate the restriction sites that define the specific haplogroups. The restriction enzymes correspond to the following code: a, *Alu*I; b, *Ava*II; c, *Dde*I; e, *Hae*III; g, *Hin*fI; i, *Msp*I; m, *Bam*HI; n, *Hae*II; p, *Bst*NI; q, *Nla*III. The horizontal branch lengths are proportional to the number of mutational events that separate the haplotypes with the exception of the 16517 *Hae*III site. In the parsimony analysis, this hypervariable site was assigned half of the weight assigned to all other sites. This tree is 192 steps in length, has consistency and retention indices of 0.768 and 0.929, respectively, and is one of the 350 MP trees generated with the TBR branch swapping algorithm. The inset illustrates the 50% majority-rule consensus of the 350 MP trees. The consensus tree is 201 steps in length and has consistency and retention indices of 0.662 and 0.880, respectively. Note that all haplogroups are retained in the consensus tree.

ally all Scandinavian mtDNAs fell into 10 mtDNA haplogroups, designated H, I, J, K, M, T, U, V, W and X. Each of these was defined by relatively ancient and stable polymorphic sites located in the coding regions of the mtDNA. An additional analysis of Tuscans from Central Italy using the diagnostic markers of these haplogroups revealed that the same haplogroups encompassed 98% of the Tuscan mtDNAs (Table 5). The predominance of these haplogroups in three divergent European populations suggests that they may encompass most of the European mtDNA variation, in a manner analogous to the A, B, C and D haplogroups that arose in Asians and encompass virtually all Native Americans (TORRONI and WALLACE 1994). Comparison of the haplogroup data from Europeans with those of Asians and Africans should permit identification of and deductions about the progenitors of Europeans.

The distribution of the 10 haplogroups in non-European populations: Previous studies have shown that the large majority of non-European mtDNAs belong to haplogroups whose distribution appears to be continent-specific (WALLACE 1995). In the Americas, haplo-

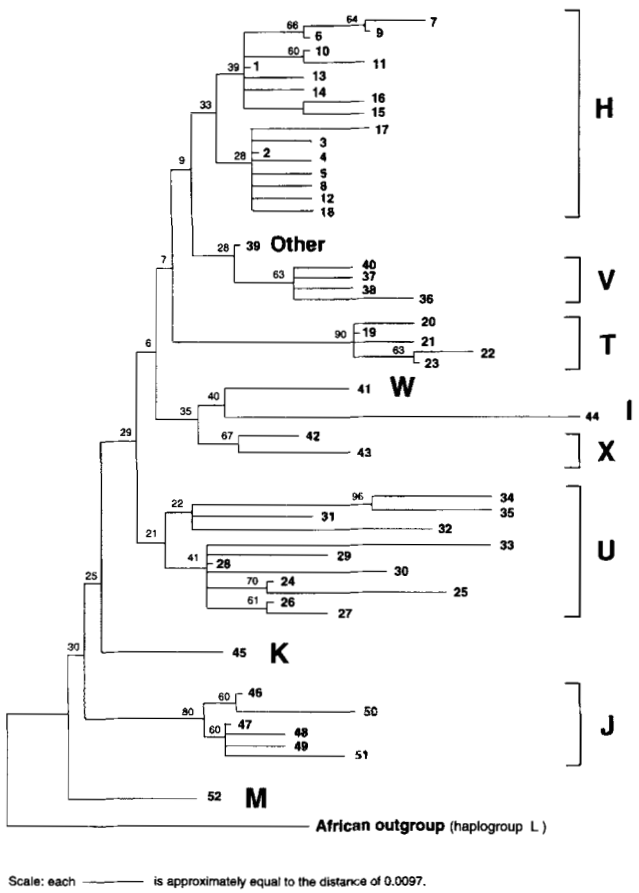

FIGURE 2.—Neighbor joining tree of Finnish and Swedish mtDNA haplotypes. This tree includes the same haplotypes of Figure 1 and was rooted by using the same African outgroup. As in Figure 1, the capital letters H, I, J, K, M, T, U, V, W and X indicate the 10 haplogroups. Numbers in bold indicate haplotypes, while the numbers at the nodes indicate bootstrap *P* values (0–100%) obtained by 500 bootstrap replications.

groups A, B, C and D encompass virtually all Native American variation; in Africa, haplogroup L represents ~70% of sub-Saharan African mtDNAs; and in Asia, haplogroup M represents ~55% of mtDNAs with other less common haplogroups (A, B, E, F, G) making up most of the remaining Asian mtDNAs (TORRONI *et al.* 1994c). With the exception of one Finnish mtDNA that belonged to haplogroup M, none of the Asian- and African-specific haplogroups were observed in our European samples. This implies that the overlap between European and non-European mtDNA variation is very limited.

A survey of published data confirms this observation. Haplogroup H, which represents ~40% of European mtDNAs, and haplogroup K, which represents ~8% of European mtDNAs, have each been observed in only three out of 1175 non-European subjects (CANN *et al.* 1987; STONEKING *et al.* 1990; BALLINGER *et al.* 1992; TORRONI *et al.* 1993a,b, 1994a,c,d; CHEN *et al.* 1995). The same survey revealed that haplogroups T, I and W have never been described in non-European populations. To date no non-European population has been screened for the enzyme *Bst*NI. As a result the frequency of the *Bst*NI np 13704 site loss of haplogroup J among Africans, Native Americans and Asians remains unknown. However, European haplotypes belonging to haplogroup J also harbor a *Hin*fI np 16065 site loss (TORRONI *et al.* 1994b) and none of the 1175 non-European mtDNAs lacked this site (CANN *et al.* 1987; STONEKING *et al.* 1990; BALLINGER *et al.* 1992; TORRONI *et al.* 1993a,b; 1994a,c,d; CHEN *et al.* 1995), indicating that haplogroup J is European-specific.

The distribution of haplogroup X in non-European populations is more difficult to define because the *Dde*I site at np 1715 appears to have arisen several independent times during human evolution, including in a few African and Asian mtDNAs. However, the combination of the *Dde*I np 1715 and np 10394 site losses has not been observed in African mtDNAs, and it has been reported in only one Asian mtDNA and one Native American population, the Ojibwa from Northern Ontario. The Ojibwa are the only Native American group showing a high proportion of mtDNAs not belonging to haplogroup A, B, C and D, and the prevalence of these additional mtDNA haplotypes, defined by the combination of the *Dde*I np 1715 and np 10394 site losses, was initially attributed to recent genetic admixure with Europeans (TORRONI *et al.* 1993a). A recent extensive reanalysis of Native American control region and haplotype data has shown that these additional Ojibwa mtDNAs belong to a well defined haplogroup (FORSTER *et al.* 1996) that corresponds to the haplogroup X observed in Europeans. However, the data from FORSTER and collaborators (1996) indicate that haplogroup X in the Ojibwa is not due to post-Columbian genetic admixture with Europeans, but it probably represents a fifth Native American founding haplogroup of Asian ancestry.

Haplogroups U and V are defined by a *Hin*fI np 12308 site gain and a *Nla*III np 4577 site loss, respectively, polymorphisms that have not been screened for in non-European populations. To assess the distribution of haplogroup U in non-Europeans, we have screened for the np 12308 mutation 116 Asians (12 Malay Chinese, 21 Vietnamese, 11 Malays, 13 Borneans, 13 Koreans, 20 Taiwanese, 26 Malay Aborigines), 51 Native Americans (2 Tlingit, 10 Dogrib, 10 Navajo, 9 Pima, 10 Maya, 10 Ticuna), and 140 Africans (60 Mandenka, 41 other Senegalese and 39 Pygmies) whose mtDNA haplotypes were previously reported (BALLINGER *et al.* 1992; TORRONI *et al.* 1992; CHEN *et al.* 1995). None of the Asian and Native American mtDNAs exhibited the 12308 mutation, though four Senegalese including one Mandenka and three Tukulors did. The Mandenka harbored haplotype AF1, while the Tukulors harbored haplotype AF2. Similar to haplogroup U mtDNAs, AF1 and AF2 lacked the *Dde*I np 10394 site.

TABLE 5

Screening of haplogroup-specific polymorphic sites in Tuscan mtDNAs

Sites[a]															
−1715 *Dde*I	+3592 *Hpa*I[b]	−4577 *Nla*III	−7025 *Alu*I	+8249 *Ava*II	−8994 *Hae*III	−9052 *Hae*II	+10028 *Alu*I	+10394 *Dde*I	+10397 *Alu*I	+12308 *Hin*fI	+13366 *Bam*HI	−13704 *Bst*NI	+15606 *Alu*I	Sample numbers	Haplogroup
−	−	−	+	−	−	−	−	−	−	−	−	−	−	3, 7, 10, 13, 16–18, 20, 24, 27, 31, 36, 39–43, 47, 49, 51	H
−	−	−	−	−	−	−	−	−	−	−	+	−	+	22, 25, 29, 32, 48	T
−	−	−	−	−	−	−	−	−	−	+	−	−	−	4, 5, 19, 30, 44	U
−	−	−	−	+	+	−	−	−	−	−	−	−	−	2	W
+	−	−	−	−	−	−	−	−	−	−	−	−	−	26, 28, 34, 38	X
+	−	−	−	+	−	−	+	+	−	−	−	−	−	9, 15	I
−	−	−	−	−	−	+	−	+	−	+	−	−	−	33, 37, 45	K
−	−	−	−	−	−	−	−	+	−	−	−	+	−	1, 11, 12, 23, 35, 46	J
−	−	−	−	−	−	−	−	+	−	−	−	+	+	21	J
−	−	−	−	−	−	−	−	−	−	−	−	−	−	14	Other

[a] The diagnostic sites are listed as either absent (−) or present (+) relative to the reference sequence (ANDERSON *et al.* 1981). A plus or minus indicates whether the listed samples harbor that polymorphism.
[b] This site characterizes the African-specific haplogroup L (CHEN *et al.* 1995).

In fact, AF2 is identical to the Finnish haplotype E28, and AF1 is closely related to the Swedish haplotype E25, with which it also shares a *Mbo*I np 3348 site gain. These findings indicate that haplogroup U, although much more prevalent in Europe, is shared between Europeans and some Sub-Saharan Africans.

To assess the distribution of haplogroup V in non-European populations, we have tested for the *Nla*III np 4577 site loss the 140 African mtDNAs previously described by CHEN *et al.* (1995), but none harbored this polymorphism. Thus haplogroup V could be absent in Africans, though its distribution in other African populations and Asians has yet to be determined.

In conclusion, it appears that at least six of the European haplogroups (haplogroups H, I, J, K, T, W) are essentially confined to European populations and probably originated after the ancestral Caucasoids became genetically separated from the ancestors of modern Africans and Asians. Haplogroup U, by contrast, is found at low frequencies in African populations.

The specificity to Caucasoids of most haplogroups observed in Europeans is supported by the age estimates of the haplogroups. These estimates (Table 6) were calculated from restriction site data from the haplotypes observed in Finns and Swedes only for those haplogroups (H, J, T, U and V) that harbored enough variation to allow these calculations. Because these estimates are based only on the mtDNA variation observed in Scandinavia, they could be biased by specific genetic phenomena associated to the peopling of this European region. However, the ages of haplogroups H and J (17,000–30,000 yr) were similar to those reported for the same haplogroups in the heterogeneous Europeans from North America (TORRONI *et al.* 1994b), and overall the relatively recent origin of Caucasoid-specific haplogroups H, J, T, and V (8000–30,000 yr) supports the hypothesis that they originated recently after the genetic and geographic separation of the ancestral Caucasoids from the ancestors of modern African and Asians. On the contrary, haplogroup U appears to be much older than the other haplogroups with an estimated age of 51,000–67,000 yr. Interestingly, this haplogroup is the only haplogroup that Europeans share with Africans. This raises the possibility that haplogroup U originated in Africa and subsequently expanded into the Middle East and Europe.

The correlation between haplogroups and control region sequences: The sequences of the control region hypervariable segments HVS-I and HVS-II for the Tuscan mtDNAs have been previously determined (FRANCALACCI *et al.* 1996). Consequently, they can be correlated with the restriction site variants of the haplogroups H, I, J, K, M, T, U, V, W, and X defined in the present study. This correlation between haplogroups and D-loop sequences had two major objectives. First, to corroborate the reliability of the haplogroups identified by haplotype and phylogenetic analyses. Indeed it is expected that, if the haplogroup classification is correct, the control region sequences within the same haplogroup should harbor unique and monophyletic sets of nucleotide polymorphisms. Second, to differentiate the control region variants that are phylogenetically associated and thus relatively ancient and stable from those that are recent or have undergone repeated forward and backward mutations.

In Table 7 the 48 control region sequences have been

TABLE 6

Sequence divergences and divergence times of haplogroups in Scandinavians

Haplogroups	No. of haplotypes	No. of subjects	Sequence divergence (%)	Divergence time (years)[a]
H	18	35	0.065	22,414–29,545
J	6	8	0.050	17,241–22,727
T	5	11	0.024	8,276–10,909
U	12	14	0.148	51,034–67,273
V	4	4	0.048	16,552–21,818

[a] The divergence times were calculated from intra-haplogroup sequence divergences using the mtDNA evolution rate of 2.2–2.9%/myr (TORRONI *et al.* 1994d).

subdivided into the observed Tuscan haplogroups (Table 5), which include all but haplogroups M and V. This subdivision confirms the validity of the haplogroup classification. Indeed, the control region sequences aggregated within the same haplogroup share a common set of ancient sequence polymorphisms (boxed in Table 7).

The haplogroup H control sequences appear to share only one common nucleotide variant, an A at residue 73 in HVS-II. Exceptions to this rule include two of the 20 haplogroup H mtDNAs (subjects 27 and 49) that harbor a G at np 73, and three nonhaplogroup H mtDNAs (subjects 4, 14 and 25) that have an A at np 73. These latter three subjects indicate that the control region sequence is highly prone to repetitive mutations. The Cambridge reference sequence is closely related to the Tuscan haplogroup H sequences and, similar to all haplogroup H mtDNAs, lacks the *Dde*I np 10394 and *Alu*I np 7025 sites. These observations suggest that the European reference sequence is also a haplogroup H mtDNA.

Four of the five Tuscan haplogroup T control region sequences are characterized by T's at nps 16294 and 16296. The fifth subject (32) has one T at np 16294. A T at np 16294 is only seen outside of haplogroup T in subject 38 of haplogroup X, while the T at np 16296 is exclusive of haplogroup T.

The haplogroup U control region sequences do not harbor any characteristic mutations, and like most haplogroups are characterized by a G at np 73. However, two subjects (19 and 44) form a subgroup defined by T's at nps 16192 and 16270.

Haplogroup W was observed in only one Tuscan sample (subject 2). Consequently, it was not possible to directly determine specific control region mutations that define this haplogroup. However, one of the 100 control region sequences obtained from British subjects (PIERCY *et al.* 1993) showed a striking homology with that of Tuscan subject 2, as both sequences share a T at np 16223, G at 73, A at 143, G at 189, C at 195, C at 204, and A at 207. Hence, these residues may be characteristic of some mtDNAs from haplogroup W.

The haplogroup X control region sequences are associated with T's at nps 16223 and 16278. The association of these two mutations is very common in African mtDNAs (GRAVEN *et al.* 1995). However, African mtDNAs lack the *Dde*I np 1715 site loss and have additional characteristic variants both in the coding and control regions.

Haplogroup I control region sequences were obtained from two Tuscan subjects. These mtDNAs shared an A at np 16129, T at 16223, C at 16311, C at 152, G at 189, C at 199, A at 203, C at 204, and C at 250 (Table 4). This same array of base substitutions has been reported in Sardinians (DI RIENZO and WILSON 1991), Britons (PIERCY *et al.* 1993) and in a heterogeneous Caucasian sample (STONEKING *et al.* 1991). Haplogroup I is also characterized by five distinctive restriction site polymorphisms. The NJ tree (Figure 2) had indicated that haplogroups I, W and X could be more closely related to each other than to other haplogroups. This suggestion appears to be partially supported by comparison of their control regions. Indeed, even though the control region sequences belonging to haplogroups I, W and X are distinguished by a substantial number of different base substitutions, they all share a T at 16223.

The haplogroup K control region sequences could be examined in three Tuscan mtDNAs and are characterized by C's at nps 16224 and 16311 of HVS-I. One of the three Tuscan mtDNAs belonging to this haplogroup lacked the C at np 16311, but this T to C transition appears to have arisen repeatedly, being also observed in both individuals from haplogroup I, as well as in some individuals from haplogroups H and J.

Haplogroup J mtDNAs are represented by seven Tuscan control region sequences and are defined by three distinct polymorphisms: a T at np 16069, a C at np 16126, and a T at np 295. The T at np 16069 is detected as a *Hin*fI site loss at np 16065 in haplotype analysis (Figure 1). The control region sequence of individual 21 is concordant with a haplogroup J classification, confirming that the *Alu*I np 15606 site found in subject 21, but normally characteristic of haplogroup T (Table 5), is due to a separate mutational event.

This correlation between nucleotide changes in the coding region and in the noncoding control region of the mtDNA permits a direct comparison of all published data obtained by both haplotype analysis and control region sequencing, thus providing additional

TABLE 7

MtDNA control region sequences from Tuscans

S[b]	H[c]	Polymorphic nucleotide positions[a]
		1111111111 1111111111 1111111111 1111111111 1111111111 11111 6666666666 6666666666 6666666666 6666666666 6666666666 66666 0001111111 1111111112 2222222222 2222222222 2222333333 33333 111111 1111112222 2222222222 33333 5690122344 5567788991 2223344456 6677788899 9999000111 22556 5667445558 8889990000 1222235569 00111 1934169457 3632669233 2340108961 3504814601 2467049168 45262 7043360230 5894593457 5256840935 99159 aa ab
Cam[d]		ACTCCTGCGC GGATCCTCCG CCTATACTCC TACGCAACCC CCCTATATAA TTTTT --CAGTCTAT GAACTTGTGG ACGTGATAAC --C-T
10	H	---------- ---------- ---------- ---------- ---------- ----- ---------- ---------- --------G- ---C-
20	H	---------- ---------- ---------- ---------- ---------- ----- ---------- ---------- --------G- ---C-
31	H	---------- ---------- ---------- ---------- ---------- ----- ---------- ---------- --------G- ---C-
41	H	---------- ---------- ---------- ---------- ---------- ----- ---------- ---------- --------G- ---C-
47	H	---------- ---------- ---------- ---------- ---------- ----- ---------- ---------- --------G- ---C-
3	H	---------- ---------- ---------- ---------- ---------- ----- ---------- ---------- --------G- C--C-
36	H	---------- ---------- ---------- ---------- ---------- ----- ---------- ---------- --------G- C--C-
39	H	---------- ---------- ---------- ---------- ---------- ----- ------T--- ---------- --------G- C--C-
16	H	------A--- ---------- ---------- ---------- --------G- ----- ------C-- ---------- --------G- ---C-
43	H	---------- ---C-----A ---------- ---------- ---------- ----- ------C-- ---------- --------G- CC-C-
51	H	---------- ---------- --------T- ---------- -------C-- --C-- ------C-- ---------- --------G- C--C-
17	H	---T------ ---------- ---------- ----T----T ---------- ----- ------C-- ---------- --------G- ---C-
18	H	---T------ ---C------ ---------- ---------- ---------- ----- ---------- ---------- --------G- CC-C-
7	H	---------- ---------- -------C-- ---------- -------C-- ----- ---------- ---T----A- --------G- C--C-
42	H	---------- -------T-- ---------T ---------- -------C-- ----- ---------- ---------- --------G- C--C-
24	H	---------- ---------- ---------- --T--G---- ---------- ----- ---------- ---------- --------G- ---C-
13	H	---------- ---------- ---------- ---------- -----C---- ----- ---------- ---------- --------G- C--C-
40	H	---------T ------C--- ---------- ---------- ---------- ---C- ---------- -----C---- -----G--G- ---C-
49	H	---------- ---------- ---------- ---------- ---------- ----- ---G--T--- -----C---- --------G- C--C-
27	H	---------- ------C--- ---------- ------GT-- ---------- ----- ---G----G- ----C----- --------G- C--C-
25	T	-----C---- ---------- ---------- -G-------- -TT--C---- ----- ---------- ---------- --------G- CC-C-
22	T	-----C---- ------C--- ---------- ---------- -TT------- ----- ---G------ ---------- --------G- C--C-
29	T	-----C---- ---------- ---------- ---------- -TT------- C---- ---G------ ---------- --------G- C--C-
48	T	----------A--------- ---------- ---------- -TT------- ----- ---G--T--- ---------- --------G- ---C-
32	T	---------- --G--TC--- ---------- ---A------ -T-------- ----- ---G---C-- ---------- --------G- ---CC
19	U	---------- -------T-- ---------- --T-----T- ---------- ----- ---G--T--- ----C----- --------G- C--C-
44	U	---------- -------T-- T-------T- --T------- ---------- ----- ---G------ -------C-- --------G- C--C-
30	U	G--------- ---------- ---------- ---------- ------G--C ----- ---G---CG- ---------- -------GG- ---C-
5	U	-------T-- ---------- ---------- C--------- ---------- ---C- ---G------ ----C----- --------G- ---C-
4	U	-----C---- ---------- ---G------ ---------- ---------- ----- -TT------- ---------- --------G- ---C-
2	W	--C------- -------T-- -T-------- ---------T T--------- -C--- ---GA--C--- --G-C--C-A --------G- ---C-
28	X	---------- ------C--- -T-------- ----T----- ---C------ ----- ---G----G- ----C----- --AC----G- C--C-
26	X	---------- ------C--- -T----T--- ----T----- ---------- C---- ---G---C-- ---------- --------G- ---C-
38	X	---------- ---------- -T-------- ----T----- -T-------- ----C ---G-C---- ---------- --------G- ---C-
34	X	----T----- ---------- -T-------- ----T----- ---------- ----- ---G------ ----C--C-A --------G- ---C-
9	I	------A--- ---C------ -T-------- ---------- -------C-- ----- ---G---C--- --G-CAC--- ------C-G- ---C-
15	I	------A--- ---C------ -T-------- ---------- -------C-- ----C ---G---C--- --G-CAC--- ------C-G- ---C-
33	K	---------- ---------- --C------- ---------- -------C-- ----- ---G---C-- ---------- --------G- ---C-
37	K	--C------- -A-------- --C--C---- ---------- -------C-- ----- ---G------ ----C----- --------G- ---C-
45	K	---------- ---------- --C------- ---------- -------[-]- ----- ---G------ ---------- --------G- ---C-

TABLE 7

Continued

Polymorphic nucleotide positions[a]

S[b]	H[c]	1111111111 6666666666 0001111111 5690122344 1934169457	1111111111 6666666666 1111111112 5567788991 3632669233	1111111111 6666666666 2222222222 2223344456 2340108961	1111111111 6666666666 2222222222 6677788899 3504814601	1111111111 6666666666 2222333333 9999000111 2467049168	11111 66666 33333 22556 45262	111111 5667445558 7043360230 aa	1111112222 8889990000 5894593457	2222222222 1222235569 5256840935	33333 00111 99159 ab
1	J	-T---C----	----T-----	----------	----------	----------	-----	---G------	----------	----A---GT	C--C-
23	J	-T---C----	----------	----------	----------	----------	-----	---G-C----	AG--------	-T--A---GT	---C-
46	J	-T---C----	----------	----------	----------	----------	-----	---G------	A---------	----A---GT	---C-
11	J	-T---C--A-	----------	----C----T	----------	----------	-----	---G--TC--	----C-----	G-------GT	--TCC
35	J	-T---C--A-	----------	T--------T	---A------	----------	-----	---G------	----------	--------GT	---C-
12	J	-T---C----	--------T-	----------	----------	-------C--	-----	---G--TC--	----------	--------GT	---C-
21	J	-T---C----	--------T-	--------T-	----------	----G-G---	-----	---G---C--	----------	--------GT	---C-
14	Other	-----C----	----------	----------	----------	----------	-----	C-T------C	----------	--------G-	CC-C-

[a] Boxes define the D-loop nucleotide variants which either by themselves, or in association with other variants, define each haplogroup. Nucleotide positions 57a, 60a, 309a, 309b indicate nucleotide insertions relative to the Cambridge sequence.
[b] S is Sample. The numbering of individual samples is according to FRANCALACCI *et al.* (1996).
[c] H is Haplogroup. Haplogroups as determined by restriction analysis screening of informative polymorphic sites.
[d] Cam is Cambridge; from ANDERSON *et al.* (1981).

information about the distribution of these haplogroups in other populations. Surveys of published control region sequences from non-European populations confirm that haplogroups I, K, J, T, W are essentially European-specific (VIGILANT *et al.* 1991; TORRONI *et al.* 1993a,b; GRAVEN *et al.* 1995). For haplogroup H, both the distinctive *Alu*I np 7025 site loss and the A residue at np 73 have only been sporadically observed in non-Europeans (STONEKING *et al.* 1991; VIGILANT *et al.* 1991; GRAVEN *et al.* 1995). As for haplogroup U, the shared African and European haplogroup, the lack of distinguishing control region variants does not permit further continental discrimination. Finally, for haplogroup X, the association between T's at at nps 16223 and 16278 is very common in African mtDNAs (GRAVEN *et al.* 1995), and thus European and African mtDNAs must be distinguished by continent-specific restriction site polymorphisms, such as the *Dde*I np 1715 site loss.

The haplogroup aggregation of control region sequences also permits identification of hypervariable nucleotides (WAKELEY 1993; COMAS *et al.* 1995). These include a C at np 16189, T at 16256, T at 150, C at 152, and C at 195, each of which appears in three or more haplogroups. A particularly high mutation rate of certain nucleotides has important implications for phylogenetic studies (BANDELT *et al.* 1995) because both maximum parsimony analysis (VIGILANT *et al.* 1989; DI RIENZO and WILSON 1991; TORRONI *et al.* 1993a,b) and pairwise difference analysis (DI RIENZO and WILSON 1991; ROGERS and HARPENDING 1992) usually assign an equal weight to all nucleotide changes.

The distribution of the 10 haplogroups in Caucasoid populations: Table 4 shows the distribution of the 10 haplogroups in the three European samples, as well as in an aggregate of Europeans from North America. Haplogroup H is the most prevalent in these samples, with a strikingly consistent frequency of ~41%. Haplogroups J, K, T and U are also very common and shared by all populations. The remaining haplogroups I, V, W and X are less common, though still shared between at least two of the three populations.

Although detailed restriction site analyses are not available for other European populations, some additional data about the distribution of specific haplogroups can be deducted from surveys that have employed a more limited number of restriction endonucleases. The linked *Bam*HI np 13366 site gain and *Msp*I np 15925 site loss of haplogroup T has been observed in 11.9% of the Greeks, 6.0% of the Albanians, 9.9% of the Sardinians, 12.6% of the Apulians, 14.6% of the Sicilians, 4.8% of the Arabs from Palestine, and 1.3% of the Ashkenazi Jews, but was not observed in the Yemenite Jews (SANTACHIARA-BENERECETTI *et al.* 1988; DE BENEDICTIS *et al.* 1989; SEMINO *et al.* 1989; TORRONI *et al.* 1990; RITTE *et al.* 1993; ASTRINIDIS and KOUVATSI 1994). These haplogroup T markers have also been seen in 3.8% of the Hindus from Northern India (SEMINO *et al.* 1991), indicating that haplogroup T is present, at varying frequencies, throughout the entire range of Caucasoid populations. A similar distribution is observed for haplogroup K defined by the *Hae*II np 9052 site loss. This polymorphism has been

described in 6.8% of the Greeks, 6.0% of the Albanians, 4.5% of the Sardinians, 9.2% of the Apulians, 12.2% of the Sicilians, 1.6% of the Arabs from Palestine, 8.0% of the Yemenite Jews, and 1.2% of the Hindus. In addition, 29.3% of the Ashkenazi Jews appear to belong to this haplogroup suggesting a recent founder effect in Central European Jewish populations (SANTACHIARA-BENERECETTI *et al.* 1988; DE BENEDICTIS *et al.* 1989; SEMINO *et al.* 1989; TORRONI *et al.* 1990; RITTE *et al.* 1993; ASTRINIDIS and KOUVATSI 1994).

Additional information about the distribution of the haplogroups in Europeans can be inferred by published studies of control region sequences, in spite of the fact that several of these studies analyzed only HVS-I (DI RIENZO and WILSON 1991; BERTRANPETIT *et al.* 1995; SAJANTILA *et al.* 1995). The specific HVS-I polymorphisms of haplogroup J, a T at np 16069 and a C at np 16126, have been observed in 12.0% of the British, 4.4% of the Basques, 5.8% of the Sardinians, 16.7% of a heterogeneous Middle Eastern sample (Bedouins, Israeli Arabs and Yemenite Jews), but none of 98 individuals from south-western India. The haplogroup K polymorphisms of C's at nps 16224 and 16226 were observed 10.0% of the British, 2.2% of the Basques, 5.8% of the Sardinians, and 2.4% of the subjects from Middle East, but were absent in the Indians. Haplogroup T mtDNAs defined by T's at nps 16294 and 16296 were observed in 10.0% of the British, 4.4% of the Basques, 13.0% of the Sardinians, and 14.3% of the individuals from Middle East, but again were not observed in Indians (DI RIENZO and WILSON 1991; PIERCY *et al.* 1993; BERTRANPETIT *et al.* 1995; MOUNTAIN *et al.* 1995).

British and Indian populations have been analyzed for both control region hypervariable segments (PIERCY *et al.* 1993; MOUNTAIN *et al.* 1995), thus providing data on the distribution of the haplogroup H np 73 A polymorphism. This data indicates that 54% of the British mtDNAs but only 1% of the Indian mtDNAs could belong to haplogroup H. This difference is statistically highly significant ($X_{1\ d.f.,Yates} = 66.6$; $P \ll 0.001$). This raises the possibility that haplogroup H frequencies may be more variable than those found in the current study.

Because most of the 10 haplogroups appear to be confined to Caucasoids, they probably originated in ancestral Caucasoid population(s) after they separated from the progenitors of modern Africans and Asians. Therefore, it is possible that frequency clines in mtDNA haplogroups might reveal the homeland and patterns of expansions of ancient proto-European populations. For instance, a steep cline appears to exist for haplogroup H, which encompasses 40–50% of mtDNAs from western European populations, but is virtually absent in Indians. Such a cline for haplogroup H has been recently confirmed by screening a number of Caucasoid populations for the absence of the 7025 *AluI* site (PASSARINO *et al.* 1996). The detection and analysis of these clines may be extremely useful in examining current models postulated to explain the bio-cultural diversity of Caucasoids (AMMERMANN and CAVALLI-SFORZA 1984; CAVALLI-SFORZA *et al.* 1988; BARBUJANI and SOKAL 1990; SOKAL *et al.* 1991; PIAZZA 1993; BARBUJANI *et al.* 1995).

Applications in studies of ancient human remains: The correlation between stable restriction site changes in the coding regions and informative polymorphisms of the control region allows haplogroup classification of control region sequences obtained from ancient human remains. Recently, the control region sequence was published from an ~5000-year-old mummified body found in the Tyrolean Alps at the border between Italy and Austria (HANDT *et al.* 1994). This sequence encompassed from np 16056 to 16409, including HSV-I. It differed from the Cambridge sequence only by a C at np 16224 and a C at np 16311. Because identical sequences were found only in Europeans, it was concluded that the mtDNA from the 5000-year-old body fitted into the genetic variation of contemporary Europeans (HANDT *et al.* 1994). The current study shows that this control region sequence is characteristic of the European-specific haplogroup K (Table 7), and therefore, that haplogroup K mtDNAs were present in the human populations living in the Alpine region ~5000 yr ago. We also know that haplogroup K is shared by all modern European populations, reaching the highest frequency values among the Ashkenazi Jews (29%) (RITTE *et al.* 1993), the Swedes analyzed in this study (14%), and the Sicilians (12%) (SEMINO *et al.* 1989). In addition, the identification of these 10 haplogroups in Europeans may permit genetic analyses of ancient human populations from Europe and the Middle East, such as the identification of the Native American-specific haplogroups A, B, C and D (TORRONI *et al.* 1992, 1993a) have allowed the classification of 50 mtDNAs from pre-Columbian Oneota of Illinois (STONE and STONEKING 1993; STONEKING 1995). Indeed, with the distinctive haplogroup restriction sites identified, mtDNAs from ancient human remains could be classified into haplogroups even when template damage permits PCR amplification of only 40- to 50-bp fragments.

Implications for studies of mtDNA diseases: One of the most important and common criterion used to determine whether a DNA mutation is causal to a disease or is a neutral polymorphism is the screening for the mutation in control populations. The pathological significance of DNA mutations is generally ruled out when the mutation is detected in the general population. However, mtDNA studies have shown that most of mtDNA variation is continent-specific, and that certain variants and haplogroups can be extremely common in certain populations and completely absent in other ethnic groups. Consequently, the application of this criterion to mtDNA disease studies requires that control populations are accurately ethnically matched to the ethnic background of the patients. However, such ethnically matched control populations are not easy to se-

lect, particularly in those regions of the World that are inhabited by populations with a heterogeneous ethnic background and have undergone through extensive genetic admixture. In recent years, a new approach has been developed in mtDNA disease study to overcome this difficulty. This method requires the comparison of the patient's mtDNA variation with that observed in the control mtDNAs that are phylogenetically most closely related, *i.e.*, with control mtDNAs belonging to the same mtDNA haplogroup observed in the patient. This approach has proven extremely valuable. For instance, it has allowed the identification of the first LHON plus dystonia mutation (14459A) in an Hispanic family from California whose mtDNA was a member of Native American haplogroup D (JUN *et al.* 1994). It has revealed that a 270-bp tandem duplication in the mtDNA control region is not associated with deletions in mitochondrial myopathies as initially suggested (BROCKINGTON *et al.* 1993), but is a polymorphism that characterizes all mtDNAs belonging to haplogroup I, one of the European-specific haplogroups (TORRONI *et al.* 1994b). It has proven that the 4336C mutation is associated with late-onset Alzheimer disease (SHOFFNER *et al.* 1993; HUTCHIN and CORTOPASSI 1995), and that this mutation arose as a single mutational event on the European-specific haplogroup H (TORRONI *et al.* 1994b). Last, it has shown that the degree of penetrance of some pathological mtDNA mutations depends on the mtDNA background on which they occur. For instance, the risk of expression of LHON in presence of the primary mutation 14484C is eightfold higher when this mutation occurs on the European-specific haplogroup J (TORRONI *et al.* 1996). These important results have been obtained in a context in which only four haplogroups (H, I, J and K) were defined among Europeans. Now, the classification of almost all European mtDNA variation in 10 haplogroups provides an additional opportunity for correlating mtDNA disease and evolutionary studies, for discriminating between neutral polymorphisms and pathological mutations, and for defining the role played by mtDNA backgrounds in disease expression.

In conclusion, this study identifies 10 mtDNA haplogroups in Europeans by haplotype comparison, phylogenetic analyses and correlation of restriction site changes and control region sequence polymorphisms. These haplogroups encompass virtually all mtDNAs observed in three European populations representative of both northwestern and southwestern Europe.

Using the combination of the informative restriction site and control region sequence polymorphisms, it has been possible to begin a survey of mtDNA haplogroup distributions throughout the Caucasoid range, suggesting that clinal distributions might exist for certain haplogroups. Moreover, the ability to discriminate among the 10 haplogroups represents a powerful tool for molecular anthropology and mtDNA disease studies. Hence, haplotype discrimination promises to continue to be a useful approach for both theoretical and applied studies of individual and global origins.

We are indebted to Dr. NILS-GORAN LARSSON for providing the DNA samples from Sweden. This work was supported by Telethon grant E.069 to A.T., by the J. Worley Brown Fellowship fund and the National Institutes of Health grant GM-46915 to D.C.W., by grants from Centro Nazionale Ricerche "Biological Archive" and the Sardinian Autonomous Region to P.F., and a grant from the Italian Ministry for Universities (MURST, 40%) to R.S.

LITERATURE CITED

AMMERMAN, A. J., and L. L. CAVALLI-SFORZA, 1984 *The Neolithic Transition and the Genetics of Populations in Europe.* Princeton University Press, Princeton, NJ.

ANDERSON, S., A. T. BANKIER, B. G. BARRELL, M. H. L. DE BRUIJN, A. R. COULSON *et al.*, 1981 Sequence and organization of the human mitochondrial genome. Nature **290:** 457–465.

ASTRINIDIS, A., and A. KOUVATSI, 1994 Mitochondrial DNA polymorphism in Northern Greece. Hum. Biol. **66:** 601–611.

BALLINGER, S. W., T. G. SCHURR, A. TORRONI, Y.-Y. GAN, J. A. HODGE *et al.*, 1992 Southeast Asian mitochondrial DNA analysis reveals genetic continuity of ancient Mongoloid migrations. Genetics **130:** 139–152.

BANDELT, H. J., P. FORSTER, B. C. SYKES and M. B. RICHARDS, 1995 Mitochondrial portraits of human populations using median networks. Genetics **141:** 743–753.

BARBUJANI, G., and S. S. SOKAL, 1990 Zones of sharp genetic change in Europe are also linguistic boundaries. Proc. Natl. Acad. Sci. USA **87:** 1816–1819.

BARBUJANI, G., G. BERTORELLE, G. CAPITANI and R. SCOZZARI, 1995 Geographical structuring in the mtDNA of Italians. Proc. Natl. Acad. Sci. USA **92:** 9171–9175.

BERTRANPETIT, J., J. SALA, F. CALAFELL, P. A. UNDERHILL, P. MORAL *et al.*, 1995 Human mitochondrial DNA variation and the origin of Basques. Ann. Hum. Genet. **59:** 63–81.

BROCKINGTON, M., M. G. SWEENEY, S. R. HAMMANS, J. A. MORGAN-HUGHES and A. HARDING, 1993 A tandem duplication in the D-loop of human mitochondrial DNAs is associated with deletions in mitochondrial myopathies. Nat. Genet. **4:** 67-71.

CANN, R. L., M. STONEKING and A. C. WILSON, 1987 Mitochondrial DNA and human evolution. Nature **325:** 31–36.

CAVALLI-SFORZA, L. L., and A. PIAZZA, 1993 Human genomic diversity in Europe: a summary of recent research and prospects for the future. Eur. J. Hum. Genet. **1:** 3–18.

CAVALLI-SFORZA, L. L., A. PIAZZA, P. MENOZZI and J. L. MOUNTAIN, 1988 Recon struction of human evolution: bringing together genetic, archaeological and linguistic data. Proc. Natl. Acad. Sci. USA **85:** 6002–6006

CAVALLI-SFORZA, L. L., P. MENOZZI and A. PIAZZA, 1994 *The History and Geography of Human Genes.* Princeton University Press, Princeton, NJ.

CHEN, Y.-S., A. TORRONI, L. EXCOFFIER, A. S. SANTACHIARA-BENERECETTI and D. C. WALLACE, 1995 Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups. Am. J. Hum. Genet. **57:** 133-149.

COMAS, D., S. PAABO and J. BERTRANPETIT, 1995 Heteroplasmy in the control region of human mitochondrial DNA. Gen. Res. **5:** 89–90.

DE BENEDICTIS, G., G. ROSE, G. PASSARINO and C. QUAGLIARELLO, 1989 Restriction fragment length polymorphism of human mitochondrial DNA in a sample population from Apulia (South Italy). Ann. Hum. Genet. **53:** 311–318.

DE LA CHAPELLE, A., 1993 Disease gene mapping in isolated populations: the example of Finland. J. Med. Genet. **30:** 857–865.

DI RIENZO, A., and A. C. WILSON, 1991 Branching pattern in the evolutionary tree for human mitochondrial DNA. Proc. Natl. Acad. Sci. USA **88:** 1597–1601.

FELSENSTEIN, J., 1993 *PHYLIP: Phylogeny Inference Package, version 3.5c.* Distributed by the author. Department of Genetics, University of Washington, Seattle.

FORSTER, P., R. HARDING, A. TORRONI and H. J. BANDELT, 1996 Ori-

gin and evolution of Native American mtDNA variation: a reappraisal. Am. J. Hum. Genet. **59:** 935–945.

FRANCALACCI, P., J. BERTRANPETIT, F. CALAFELL and P. UNDERHILL, 1996 Sequence diversity of the control region of mitochondrial DNA in Tuscany and its implications for the peopling of Europe. Am. J. Phys. Anthropol. **100:** 443–460.

GILES, R. E., H. BLANC, H. M. CANN and D. C. WALLACE, 1980 Maternal inheritance of human mitochondrial DNA. Proc. Natl. Acad. Sci. USA **77:** 6715–6719.

GIMBUTAS, M., 1991 *The Civilization of the Goddess.* Harper, San Francisco.

GRAVEN, L., G. PASSARINO, O. SEMINO, P. BOURSOT, A. S. SANTACHIARA-BENERECETTI *et al.*, 1995 Evolutionary correlation between control region sequence and restriction polymorphisms in the mitochondrial genome of a large Senegalese Mandenka population. Mol. Biol. Evol. **12:** 334–345.

HANDT, O., M. RICHARDS, M. TROMMSDORFF, C. KILGER, J. SIMANAINEN *et al.*, 1994 Molecular genetic analyses of the Tyrolean Ice Man. Science **264:** 1775–1778.

HEDGES, S. B., S. KUMAR, K. TAMURA and M. STONEKING, 1991 Human origins and analysis of mitochondrial DNA sequences. Science **255:** 737–739.

HORAI, S., R. KONDO, Y. NAKAGAWA-HATTORI, S. HAYASHI, S. SONODA *et al.*, 1993 Peopling of the Americas, founded by four major lineages of mitochondrial DNA. Mol. Biol. Evol. **10:** 23–47.

HUTCHIN, T., and G. CORTOPASSI, 1995 A mitochondrial DNA clone is associated with increased risk for Alzheimer disease. Proc. Natl. Acad. Sci. USA **92:** 6892–6895.

JUN, A. S., M. D. BROWN and D. C. WALLACE, 1994 A mitochondrial DNA mutation at nucleotide pair 14459 of the NADH dehydrogenase subunit 6 gene associated with maternally inherited Leber hereditary optic neuropathy and dystonia. Proc. Natl. Acad. Sci. USA **91:** 6206–6210.

KUMAR, S., K. TAMURA and M. NEI, 1993 *MEGA: Molecular Evolutionary Genetics Analysis, version 1.01.* The Pennsylvania State University, University Park, PA.

LAUBER, J., C. MARSAC, B. KADENBACH and P. SEIBEL, 1991 Mutations in mitochondrial tRNA genes: a frequent cause of neuromuscular diseases. Nucleic Acids Res. **19:** 1393–1397.

LI, W.-H., and A. ZHARKIKH, 1995 Statistical tests of DNA phylogenies. Syst. Biol. **44:** 49–63.

MIYATA, T., H. HAYASHIDA, R. KIKUNO, M. HASEGAWA, M. KOBAYASHI *et al.*, 1982 Molecular clock of silent substitution: at least six-fold preponderance of silent changes in mitochondrial genes over those in nuclear genes. J. Mol. Evol. **19:** 28–35.

MOUNTAIN, J. L., J. M. HEBERT, S. BHATTACHARYYA, P. A. UNDERHILL, C. OTTOLENGHI *et al.*, 1994 Demographic history of India and mtDNA-sequence diversity. Am. J. Hum. Genet. **56:** 979–992.

NEI, M., and F. TAJIMA, 1983 Maximum likelihood estimation of the number of nucleotide substitutions from restriction site data. Genetics **105:** 207–217.

PASSARINO, G., O. SEMINO, L. F. BERNINI and A. S. SANTACHIARA-BENERECETTI, 1996 Pre-Caucasoid and Caucasoid genetic features of Indian population revealed by mtDNA polymorphisms. Am. J. Hum. Genet. **59:** 927–934.

PIAZZA, A., 1993 Who are the Europeans? Science **260:** 1767–1769.

PIERCY, R., K. M. SULLIVAN, N. BENSON and P. GILL 1993 The application of mitochondrial DNA typing to the study of white Caucasian genetic identification. Int. J. Leg. Med. **106:** 85–90.

RITTE, U., E. NEUFELD, F. M. PREGER, M. GROSS, I. HAKIM *et al.*, 1993 Mitochondrial DNA affinities of several Jewish communities. Hum. Biol. **65:** 359–385.

ROGERS, A. R., and H. HARPENDING, 1992 Population growth make waves in the distribution of pairwise genetic distances. Mol. Biol. Evol. **9:** 552–569.

SAJANTILA, A., P. LAHERMO, T. ANTINNEN, M. LUKKA, P. SISTONEN *et al.*, 1995 Genes and languages in Europe: an analysis of mitochondrial lineages. Gen. Res. **5:** 42–52.

SANTACHIARA-BENERECETTI, A. S., R. SCOZZARI, O. SEMINO, A. TORRONI, A. BREGA *et al.*, 1988 Mitochondrial DNA polymorphisms in Italy. II. Molecular analysis of new and rare morphs from Sardinia and Rome. Ann. Hum. Genet. **52:** 39-56.

SCHURR, T. G., S. W. BALLINGER, Y. Y. GAN, J. A. HODGE, D. A. MERRIWETHER *et al.*, 1990 Amerindian mitochondrial DNAs have rare Asian mutations at high frequencies, suggesting they derived from four primary maternal lineages. Am. J. Hum. Genet. **46:** 613–623.

SEMINO, O., A. TORRONI, R. SCOZZARI, A. BREGA, G. DE BENEDICTIS *et al.*, 1989 Mitochondrial DNA polymorphisms in Italy. III. Population data from Sicily: a possible quantitation of maternal African anacestry. Ann. Hum. Genet. **53:** 193–202.

SEMINO, O., A. TORRONI, R. SCOZZARI, A. BREGA and A. S. SANTACHIARA-BENERECETTI, 1991 Mitochondrial DNA polymorphisms among Hindus: a comparison with the Tharu of Nepal. Ann. Hum. Genet. **55:** 123–136.

SHOFFNER, J. M., M. D. BROWN, A. TORRONI, M. T. LOTT, M. CABELL *et al.*, 1993 Mitochondrial DNA variants observed in Alzheimer and Parkinson disease patients. Genomics **17:** 171–184.

SOKAL, R. R., N. L. ODEN and C. WILSON, 1991 Genetic evidence for the spread of agriculture in Europe by demic diffusion. Nature **351:** 143–145.

SOODYALL, H., L. VIGILANT, A. V. HILL, M. STONEKING and T. JENKINS, 1996 MtDNA control-region sequence variation suggests multiple independent origins of an "Asian-specific" 9-bp deletion in sub-saharan Africans. Am. J. Hum. Genet. **58:** 595-608.

STONE, A. C., and M. STONEKING, 1993 Ancient DNA from a Pre-Columbian Amerindian population. Am. J. Phys. Anthropol. **92:** 463–471.

STONEKING, M., 1995 Ancient DNA: how do you know when you have it and what can you do with it? Am. J. Hum. Genet. **57:** 1259–1262.

STONEKING, M., L. B. JORDE, K. BHATIA and A. C. WILSON, 1990 Geographic variation in human mitochondrial DNA from Papua New Guinea. Genetics **124:** 717-733.

STONEKING, M., D. HEDGECOCK, R. G. HIGUCHI, L. VIGILANT and H. A. ERLICH, 1991 Population variation of human mtDNA control region sequences detected by enzymatic amplification and sequence-specific oligonucleotide probes. Am. J. Hum. Genet. **48:** 370–382.

SWOFFORD, D., 1993 *Phylogenetic Analysis Using Parsimony (PAUP), Version 3.1.1.* Illinois Natural History Survey, Champaign, IL.

TORRONI, A., and D. C. WALLACE, 1994 MtDNA variation in human populations and implications for detection of mitochondrial DNA mutations of pathological significance. J. Bioenerg. Biomemb. **26:** 261–271.

TORRONI, A., O. SEMINO, G. ROSE, G. DE BENEDICTIS, C. BRANCATI *et al.*, 1990 Mitochondrial DNA polymorphisms in the Albanian population of Calabria (Southern Italy). Int. J. Anthropol. **5(2):** 97–104.

TORRONI, A., T. G. SCHURR, C.-C. YANG, E. J. E. SZATHMARY, R. C. WILLIAMS *et al.*, 1992 Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migrations. Genetics **130:** 153–162.

TORRONI, A., T. G. SCHURR, M. F. CABELL, M. D. BROWN, J. V. NEEL *et al.*, 1993a Asian affinities and continental radiation of the four founding Native American mtDNAs. Am. J. Hum. Genet. **53:** 563–590.

TORRONI, A., R. I. SUKERNIK, T. G. SCHURR, Y. B. STARIKOVSKAYA, M. F. CABELL *et al.*, 1993b MtDNA variation of Aboriginal Siberians reveals distinct genetic affinities with Native Americans. Am. J. Hum. Genet. **53:** 591–608.

TORRONI, A., Y.-S. CHEN, O. SEMINO, A. S. SANTACHIARA-BENECERETTI, R. C. SCOTT *et al.*, 1994a MtDNA and Y-chromosome polymorphisms in four Native American populations from Southern Mexico. Am. J. Hum. Genet. **54:** 303–318.

TORRONI, A., M. T. LOTT, M. F. CABELL, Y.-S. CHEN, L. LAVERGNE *et al.*, 1994b MtDNA and the origin of Caucasians. Identification of ancient Caucasian-specific haplogroups, one of which is prone to a recurrent somatic duplication in the D-loop region. Am. J. Hum. Genet. **55:** 760–776.

TORRONI, A., J. A. MILLER, L. G. MOORE, S. ZAMUDIO, J. ZHUANG *et al.*, 1994c Mitochondrial DNA analysis in Tibet. Implications for the origin of the Tibetan population and its adaptation to high altitude. Am. J. Phys. Anthropol. **93:** 189–199.

TORRONI, A., J. V. NEEL, R. BARRANTES, T. G. SCHURR and D. C. WALLACE, 1994d A mitochondrial DNA "Clock" for the Amerinds and its implications for timing their entry into North America. Proc. Natl. Acad. Sci. USA **91:** 1158–1162.

TORRONI, A., V. CARELLI , M. PETROZZI, M. TERRACINA, P. BARBONI *et al.*, 1996 Detection of the mtDNA 14484 mutation on an African-specific haplotype. Implications about its role in causing

Leber hereditary optic neuropathy. Am. J. Hum. Genet. **59:** 248–252.

VIGILANT, L., R. PENNINGTON, H. HARPENDING, T. D. KOCHER and A. C. WILSON, 1989 Mitochondrial DNA sequences in single hairs from a southern African population. Proc. Natl. Acad. Sci. USA **86:** 9350–9354.

VIGILANT, L., M. STONEKING, H. HARPENDING, K. HAWKES and A. C. WILSON, 1991 African populations and the evolution of human mitochondrial DNA. Science **253:** 1503–1507.

WAKELEY, J., 1993 Substitution rate variation among sites in hypervariable region I of human mitochondrial DNA. J. Mol. Evol. **37:** 613–623.

WALLACE, D. C., 1995 Mitochondrial DNA variation in human evolution, degenerative disease, and aging. Am. J. Hum. Genet. **57:** 201–223.

WALLACE, D. C., J. H. YE, S. N. NECKELMANN, G. SINGH, K. A. WEBSTER *et al.*, 1987 Sequence analysis of cDNAs for the human and bovine ATP synthase beta subunit: mitochondrial DNA genes substain seventeen times more mutations. Curr. Genet. **12:** 81–90.

WARD, R. H., B. L. FRAZIER, K. DEW-JAGER and S. PAABO, 1991 Extensive mitochondrial diversity within a single Amerindian tribe. Proc. Natl. Acad. Sci. USA **88:** 8720–8724.

Communicating editor: W.-H. LI

APPENDIX A

Polymorphic sites that define Finnish and Swedish mtDNA haplotypes

Polymorphic sites	Haplotypes
	1111111111222222222233333333334444444444555
	1234567890123456789012345678901234567890123456789012
259a	000000000000000000000000000000010000000000000000000
322e	111111111111111111111111111111101111111111111111111
679c	111111111111111111111111111101111111111111111111111
740j	111111111111111111111111111111111111111111111111110
853c	111111111111111111111111111111101111111111111111111
951j	111111111101111111111111111111111111111111111111111
1338a	000000000000000000000000010000000000000000000000000
1715c	111111111111111111111111111111111111111100011111111
1718a	000000000000000000000000001000000000000000000000000
1842k	001000000000000000000000000000000000000000000000000
1941f	000000000000000000000000000000010000000000000000000
2293a	000000000000000000000000000001000000000000000000000
2831g	000000000000000000000000000000000000000010000000000
3337k	111111111111111111111101111111111111111111111111111
3348j	000000000000000000000010000000000000000000000000000
3391e	000000000000000000000000000000000000000000000001000
4332b	000000000000001000000000000000000000000000000000000
4529n	111111111111111111111111111111111111111111101111111
4577q	111111111111111111111111111111111100101111111111111
4643k	000000000000000000000000000001000000000000000000000
4769a	000000000110000000000000000000000000000000000000000
4990a	111111111111111111111111111111101111111111111111111
5003c	111110010111111111111111111111111111111111111111111
5261e	111111111111111111111111111111111101111111111111111
5315e	000000000000000000000000000000000000000001000000000
6260e	111111111111111110111111111111111111111111111111111
7025a	000000000000000001111111111111111111111111111111111
7474a	111111111111111111111111111111111111111111111111101
8165e	000000000000000000000000000010000000000000000000000
8249b/8250e	000000000000000000000000000000000000000100100000000
8268a	000010000000000000000000000000000000000000000000000
8838e	111111111111111111111111111111111111101111111111111
8858f*	111111111111111111111111111111111111111111111111111
8901e	000000000000000000000000000000000000000001000000000
8921f	000000000000000000000000000000000000000000000000100
8994e	111111111111111111111111111111111111111101111111111
9025e	111111111111111111111111111111111111111111111101111
9052n/9053f	111111111111111111111111111111111111111111101111111
9209g	000000000000000000000000000000000100000000000000000
9266e	111111111111111101111111111111101111111111111111111
9380f	111111111111101111111111111111111111111111111111111
9612k	000000000000000000000000000000010000000000000000000
9644a	111111111111011111111111111111111111111111111111111
10028a	000000000000000000000000000000000000000000100000000

TABLE APPENDIX A

Continued

Polimorphic sites	Haplotypes
10097a	00010000000000000000000000000000000000000000000000000
10394c	00000000000000000000000000000000000000000000111111111
10397a	00000000000000000000000000000000000000000000000000001
10746c	00000000000000000000010000000000000000000000000000000
11196k	00000000000000000000000000000000010000000000000000000
12308g	00000000000000000000001111111111110000000010000000
12406h	11111111111111111111011111111111111111111111111111111
12679j	00000000000000000000000000000000000000010000000000
12891c	11111111111111111111111111111111111111111111111111011
13103g/13104j	11111111111111111111111111111110111111111111111111
13325k	11111101111111111111111111111111111111111111111111111
13366m/13367i/13367b	00000000000000000111110000000000000000000000000000
13651	00000000000000000000000000000100000000000000000000
13702e*	00000000000000000000000000000000000000000000000000
13704p	11111111111111111111111111111111111111111111110000001
13730g	00000000000000000000000000000001100000000000000000
13968j	00000000000000000000000000000010000000000000000000
14015a	11111111111111111111111111111111111111101111111111
14068l	00000000000000000000000000000010000000000000000000
14199o*	00000000000000000000000000000000000000000000000000
14268g*	11111111111111111111111111111111111111111111111111
14347k	00000010100000000000000000000000000000000000000000
14368g*	00000000000000000000000000000000000000000000000000
14493c	00000000000000010000000000000000000000000000000000
14855c	00000000000000000000000000000100000000000000000000
15172e	11111111111111111111111111111111101111111111111111
15254r	11111111111111111111111111111111111111111111111101
15606a	00000000000000000111100000000000000000000000000000
15754c	00000000000000000000000000000000000000000100000000
15872k	00000000000000000000000000000001100000000000000000
15882b/15883e	00000010000000000000000000000000000000000000000000
15925i	11111111111111111000001111111111111111111111111111
16049k	11111111111111110111111111111111111111111111111111
16065g	11111111111111111111111111111111111111111111110000001
16143l	00000000000000000000000011000000000000000000000000
16145e	00000000000000000000000000010000000000000000000000
16170j	00000000001000000000000000000000000000000000000000
16215j	00000000000000010000000000000000000000000000000000
16224l	00000000000000000000000010000000000000000000000000
16303k	11111111111110011111001111111111111111111111111111
16389g/16390b	00000000000000000000000000000100000000000000000000
16389m/16390j/16390b	00000000000000000000000000000000000000000100000000
16398e	00000000000000000000001100000000000000000000000000
16453i	11111111111111111111111111111110011111111111111111
16517e	01111001000100001111111000000011011000001111111000101

The 84 polymorphic sites that define 52 distinct mtDNA haplotypes (E1–E52) observed in 49 Finns and 37 Swedes. A "1" indicates the presence of a site and a "0" indicates the absence of a site. Sites are numbered from the first nucleotide of the recognition sequence according to the published sequence (ANDERSON *et al.* 1981). The restriction enzymes used in the analysis are designated by the following single-letter code: a, *Alu*I; b, *Ava*II; c, *Dde*I; e, *Hae*III; f, *Hha*I; g, *Hin*fI; h, *Hpa*I; i, *Hpa*II; j, *Mbo*I; k, *Rsa*I; l, *Taq*I; m, *Bam*HI; n, *Hae*II; o, *Hinc*II; p, *Bst*NI; q, *Nla*III; r, *Acc*I. Sites separated by a diagonal line indicate either simultaneous site gains or site losses for two different enzymes or a site gain for one enzyme and a site loss for another because of a single inferred nucleotide substitution. Such sites are considered to be only one restriction site polymorphism in the parsimony analysis. Sites marked with an asterisk were found to be present or absent in all samples contrary to the published sequence.